PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/00, C12P 21/00, A01H 1/04

A1

(11) International Publication Number: **WO 95/21249**

(43) International Publication Date: 10 August 1995 (10.08.95)

(21) International Application Number: PCT/US95/01495

(22) International Filing Date: 3 February 1995 (03.02.95)

(30) Priority Data:
08/192,152 3 February 1994 (03.02.94) US

(71) Applicant: THE SCRIPPS RESEARCH INSTITUTE [US/US]; 10666 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors: BEACHY, Roger, N.; 751 Caminito Bassano, La Jolla, CA 92037 (US). MARCOS, Jose, F.; 4249 Nobel Drive, No. 28, San Digeo, CA 92122 (US).

(74) Agents: BOSTICH, June, M. et al.; Spensley Horn Jubas & Lubitz, 1880 Century Park East, 5th floor, Los Angeles, CA 90067 (US).

(81) Designated States: AU, CA, FI, JP, NO, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: A CASSETTE TO ACCUMULATE MULTIPLE PROTEINS THROUGH SYNTHESIS OF A SELF-PROCESSING POLYPEPTIDE

(57) Abstract

A cassette for simultaneous expression of two or more heterogenous peptides in equimolar amounts and based upon the nuclear inclusion (NIa) protease from tobacco etch potyvirus. The heterogenous peptides are translated and incorporated into a polypeptide that also includes the protease which releases the heterologous proteins post translationally by autoproteolytic reaction.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CS	Czechoslovakia
CZ	Czech Republic
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgystan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakhstan
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SI	Slovenia
SK	Slovakia
SN	Senegal
TD	Chad
TG	Togo
TJ	Tajikistan
TT	Trinidad and Tobago
UA	Ukraine
US	United States of America
UZ	Uzbekistan
VN	Viet Nam

A CASSETTE TO ACCUMULATE MULTIPLE PROTEINS THROUGH SYNTHESIS OF A SELF-PROCESSING POLYPEPTIDE

This invention was made with government support under Grant Nos. RO1-Al 27161-05A1 from the National Institutes of Health. The government has certain rights in this invention.

BACKGROUND OF THE INVENTION

1. *Field of the Invention*

This invention relates to methods for plant transformation to enhance and control gene expression. More particularly, this invention relates to a method for expressing more than one transgenic gene in plants in equimolar amounts from a single promoter.

2. *Description of Related Art*

In recent years, development of plant transformation techniques and strategies for enhancing and controlling gene expression have broadened the practical applications of plant biotechnology. However, the potential of all these techniques must deal with the problems encountered when more than one transgene is expressed *in planta*.

Current approaches to expressing more than one gene in transgenic plants require the use of multiple promoters, which in itself presents problems related to levels of expression from each promoter. For example, the relative levels of expression in potato plants of two genes encoding two viral coat proteins (CP), which were introduced via a single Ti-derived transformation vector, were different in different plant lines (C. Lawson, *et al.*, *Bio/Technology*, 8:127-134, 1990). In an alternative approach, plants are retransformed with a second gene, but this technique may induce gene silencing effects (M. Matzke, *et*

al., *EMBO J.*, 8:643-649, 1989; T. Fujiwara, *et al.*, *Plant Cell Rep.*, 12:133-138, 1993). In addition, sexual crossing of different transgenic lines may enhance or inhibit gene expression depending on gene copy number and the nature of the gene insertion (S. Hobbs, *et al.*, *Plant Mol Biol.*, 21:17-26, 1993). Therefore, relative levels of expression of two transgenes in a plant cannot be predicted with the use of any of these different approaches, and rather are a consequence of experimental variability.

Therefore, an alterative mechanism to express multiple genes in a single transgenic line, for instance in techniques designed to improve pathogen-derived protection against plant viruses is desirable. Systems which allow equimolar accumulation of two or more proteins under the control of a single transcriptional promoter, would avoid the problems outlined above, while providing the additional advantages of producing equal amounts of the two transgenes in each plant.

Several plant and animal viruses encode proteinases that cleave viral polypeptides yielding mature proteins. For instance, plant potyviral genomes are expressed through the translation of a single polypeptide which is processed to release multiple individual viral proteins (J. Riechmann, *et al.*, *J. Gen. Virol.*, 73:1-16, 1992). Three viral proteinase activities have been implicated in this processing (J. Carrington, *et al.*, *EMBO J.*, 9:1347-1353, 1990; J. Verchot, *et al.*, *Virology*, 185:527-535, 1991). One of these, associated with the nuclear inclusion (NIa) protein, has been widely studied in the case of tobacco etch potyvirus (TEV) (J. Carrington, *et al.*, *J. Virol.*, 62:2313-2320, 1988; J. Carrington, *et al.*, *J. Virol.*, 61:2540-2548, 1987), and is responsible for several processing events involving the large viral polypeptide. NIa from TEV functions during post-translational processing through the recognition and cleavage of a specific heptapeptide (J. Carrington, *et al.*, *Proc. Nat. Acad. Sci. USA*, 85:3391-3395, 1988; W. Dougherty, *et al.*, *EMBO J.*, 7:1281-1287, 1988). Taking advantage of this well-characterized proteinase activity, an expression cassette based on the TEV-NIa protein has been developed. This cassette

vector allows the synthesis of two or more proteins in equimolar amounts as part of a polyprotein that is cleaved into individual mature proteins by the NIa proteolytic activity.

SUMMARY OF THE INVENTION

A cassette expression vector based on the nuclear inclusion (NIa) protease from tobacco etch virus (TEV) allows the transcription and translation of a nucleotide sequence comprising the TEV NIa coding region flanked on each side by its heptapeptide cleavage sequences and insertion sites for in frame insertion of two different open reading frames coding for heterologous proteins. Upon translation, of the resulting polypeptide the protease releases the two heterologous proteins in equimolar amounts by autoproteolytic reaction. Therefore, the invention provides a method for obtaining equimolar amounts of different proteins expressed under the control of a common promoter. Alternatively, a plurality of insertions sites can be engineered into a cassette containing a single TEV NIa protease gene for production of a plurality of peptides. *In vitro* or *in vivo*, the expression cassette functions to express genes encoding two or more different heterogeneous peptides from a single polypeptide by post translational self-cleavage by the NIa protease.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A is a schematic diagram of a TEV-NIa-based expression cassette vector pPR01. The open box represents the NIa open reading frame. The shaded areas enlarged above show (as both nucleotide and amino acid sequence) the heptapeptide recognition sequence for the NIa proteolytic activity at both N- and C-termini of NIa; the engineered *Sma I* and *Stu I* cloning sites (underlined) for the in frame introduction of different genes; and start ATG and stop TGA codons. The NIa processing site between Gln and Gly is indicated as an open arrowhead. The sequence of the TEV 5' non-translated region is also indicated by a black arrow upstream of the NIa coding sequence. Relevant unique restriction enzyme sites are indicated: Ba (*BamH* I), Bg (*BgI* II), Ec (*EcoR* I), Sa (*SaI* I), Sc (*Sac* I), Sm (*Sma* I), and St (*Stu* I).

Figure 1B is a detailed restriction map of pPRO1 displaying the nucleotide sequence and the amino acid sequence of the NIa protease (SEQUENCE I.D. NO. 6).

Figure 1C is a schematic diagram showing amino acid additions that result at N- and C-termini of proteins cloned at the *Sma* I or *Stu* I enzyme restriction insertion sites of expression vector pPR01 upon translation and subsequent proteolytic processing. The amino acid represented by X depends upon the particular restriction site used for cloning and can be coincident with amino acids in the cloned proteins in some cases.

Figure 2 shows an autoradiograph of an SDS-PAGE gel indicating the results of *in vitro* translation of RNA transcribed from the pPRO1 expression cassette. Translation reactions were programmed with 1 μg of brome mosaic virus (BMV) RNAs (lane B), with no RNA added (lane 0), and with RNA transcribed *in vitro* from pPR01 (lane 1). The molecular mass (in kDa), positions of the major proteins translated from BMV RNAs, and the position of the 49 kDa TEV NIa protein are indicated.

Figure 3A shows a schematic representation of six different polypeptides translated transcribed *in vitro* from different pPRO1-derived constructs containing the TMV CP sequence. Open boxes represent the TEV-NIa sequence. Striped boxes represent the TMV CP sequence contained in the insertion site. The names of the constructs and the expected molecular mass of the translated and processed products are indicated. Q/G indicates the amino acid residues at the cleavage sequence in constructs cloned in pPRO1; whereas H/G indicates the His to Gln mutation at -1 position that inhibits processing by NIa in constructs cloned in pPRO4.

Figure 3B shows an autoradiograph of an SDS-PAGE gel containing *in vitro* translation products obtained from the constructs shown in Figure 3A.
The vertical axis and lane assignments are the same as described for Figure 3C below.

Figure 3C shows fluorographs of immunoprecipitation analyses using anti-TMV CP antibody with aliquots from the translation samples shown in Figure 3B. In Figures 3B and 3C, translation reactions were programmed with no RNA added (lane 0); with RNA transcribed *in vitro* from pPRO1 (lane 1); pPRO1.NT (lane 2); pPRO1.TN (lane 3); pPRO1.TΔN (lane 4); pPRO4.NT (lane 5); and pPRO4.TN (lane 6). The molecular mass (in kDa) and positions of ^{14}C-labeled protein markers are indicated. T= TMV coat protein; N = NIa protease

Figure 4 shows the results of *in vitro* translation of RNAs transcribed from pPRO1 constructs containing TMV CP and SMV CP coding sequences inserted at two sites in the cassette.

Figure 4A is a schematic diagram representing the vectors pPRO1.SNT and pPRO1.TNS. The open box represents the TEV-NIa sequence. Striped and dotted boxes represent TMV CP and SMV CP sequences, respectively that have been inserted into the cassette insertion sites. S = SMV coat protein.

Figure 4B shows an autoradiograph of an SDS-PAGE gel with *in vitro* translation products obtained from pPRO1.SNT and pPRO1.TNS vectors. Translation reactions were programmed with no RNA added (lane 0); with RNA transcribed *in vitro* from pPRO1 (lane 1); pPRO1.SNT (lane 2); and pPRO1.TNS (lane 3). The molecular mass (in kDa), positions of the major proteins translated from BMV RNAs, and the positions of the TEV NIa, SMV CP and TMV CP are indicated.

Figure 5 shows the results of *in vitro* translation of RNAs transcribed from a pPRO1 vector containing SMV CP and uidA (β-glucuronidase, GUS) coding sequences in the two insertion sites.

Figure 5A shows a schematic diagram representing the vector pPRO1.SNG. The open box represents TEV-NIa sequences. Dotted and striped boxed represent SMV CP and uidA (β-glucuronidase) sequences, respectively.
G = uidA, GUS enzyme.

Figure 5B shows an autoradiograph of an SDS-PAGE gel with *in vitro* translation products obtained from cassette vector pPRO1.SNG. Positions of TEV NIa, GUS, and SMV CP proteins are indicated. Translation reactions were programmed with no RNA added (lane 0); and with RNA transcribed *in vitro* either from pPRO1 (lane 1); or pPRO1.SNG (lane 2). Molecular mass (indicated in kDa), and positions of proteins translated from BMV RNAs is indicated: TEV NIa, GUS, and SMV CP proteins are also indicated. A black arrowhead indicates the position of a 110 kDa polypeptide present in small amounts.

Figure 5C shows a photograph of an SDS PAGE gel used in a time course *in vitro* translation reaction with vector pPRO1.SNG. Samples were withdrawn at times (in minutes) indicated at the top of each lane. At an incubation time of 15 minutes on SDS-PAGE, no 149 kDa precursor polypeptide could be detected.

DETAILED DESCRIPTION OF THE INVENTION

In TEV, the NIa protease is synthesized as part of the polyprotein that results from the translation of the TEV genome. The genomic sequence of TEV, first disclosed by R. Allison, *et al. (Virology,* 154:9-20, 1986) is publicly available from EMBL and Genebank database under accession number M15239. NIa recognizes and cleaves specific sequences of seven amino acids (heptapeptide) contained in the polyprotein and is responsible for partial processing of the viral polyprotein. Heptapeptide cleavage sequences recognized by the NIa from TEV (immediately 5-prime and 3-prime) have been shown to be Glu-X-X-Tyr-X-Gln-Gly (SEQUENCE I.D. NO. 1) or Glu-X-X-Tyr-X-Gln-Ser (SEQUENCE I.D. NO. 2) wherein X can be any amino acid (J. Carrington, *et al.*, 1988, *supra* and W. Dougherty, *et al., supra*). Cleavage location by TEV-NIa protease is after the Glu amino acid. In one embodiment of the present invention, the self-recognized cleavage sequence at the N-terminini of the NIa protease is Glu-Pro-Val-Tyr-Phe-Gln-Gly (SEQUENCE I.D. NO. 3) and the self-recognized cleavage sequence at the C-termini is Glu-Leu-Val-Tyr-Ser-Gln-Gly (SEQUENCE I.D. NO. 4). These two heptapeptides are the ones that bracket the NIa protein in the TEV polyprotein.

NIa releases itself from the polyprotein in an autoproteolytic reaction attacking at the cleavage sequences (J. Carrington, *et al., Virology,* 160:355-362, 1987), and is active both *in cis*, processing polypeptides in which it is included, and *in trans*, simultaneously cleaving different polypeptides. The *cis* protease activity of NIa has been assayed with different TEV polyproteins produced *in vitro* which contained NIa and either naturally occurring or mutated versions of the cleavage sequence (J. Carrington, *et al., J. Virology,* 1988, 1987, *supra*). Protease activity *in trans* has been observed in many studies using as substrates TEV polyproteins that were labeled *in vitro* and incubated with NIa extracted from infected plants.

The TEV-NIa based expression cassette provided herein has been constructed to exploit the protease activity of NIa in a self-processing polypeptide in order to express two or more different proteins in equimolar amounts. For instance, cassette vector, named pPRO1, shown in Figure 1, was obtained by PCR amplification using as template a full length TEV cloned cDNA. It comprises PRO1 (SEQUENCE ID NO. 5), which includes an open reading frame encompassing the NIa sequence (TEV nucleotides 5673 to 6983 as numbered in R. Allison, *et al., Virology*, 154:9-20, 1986) as well as the target heptapeptides located at its N-terminus (SEQUENCE ID NO. 3) and C-terminus (SEQUENCE ID NO. 4). The TEV-NIa based cassette described herein also provides at least two blunt end restriction sites, preferably unique, that allow the in frame insertion of heterologous protein sequences vector for expression as part of a self-processing polypeptide. As used herein the term "heterologous" shall have the meaning that the gene inserted into the cassette insertion site is not native to TEV.

For instance, in pPRO1 one insertion site is provided by a *Sma I* restriction enzyme site at the N-terminus of the TEV NIa sequence, and the other insertion site is provided by a *Stu* I restriction enzyme site at the C-terminus. In addition, the cassette optionally provides a start codon, preferably ATG, and a stop codon, preferably TGA, engineered upstream of the 5-prime site and downstream of the 3-prime site, respectively. For instance, in vector pPRO1, which provides two insertion sites, an ATG start codon is upstream of the *Sma* I site, and a TGA stop codon is downstream of the *Stu* I site. In addition, the TEV-NIa based vectors herein preferably include upstream of the open reading frame the 144 nucleotide 5' non-translated region from TEV RNA, which has been shown to enhance translation *in vitro* and *in vivo* (J. Carrington and D. Freed, *J. Virol*, 64:1590-1597, 1990).

One skilled in the art will appreciate that the techniques described herein could be used to insert more than two unique restriction endonuclease sites and heptapeptide recognition

sequences into the expression cassette, so as to express more than two heterologous proteins. Thus, the number of foreign proteins translated as part of a NIa-containing polyprotein is not, theoretically, limited to two, and embodiments of the cassette vector are contemplated within the scope of this invention wherein more than two insertion sites are useful for simultaneous expression of more than two proteins in equimolar amounts. In the embodiment of the invention utilizing more than one restriction site on one or both sides of the gene encoding the NIa protease and its flanking self-recognition sequences, it will be necessary to provide additional NIa protease self-recognition sequences between adjacent recognition sequences to allow for post translational self-cleavage by the NIa protease. A single protease is sufficient to cleave multiple sites within the single polypeptide produced from expression of the cassette.

PRO1 (Figure 1B; SEQUENCE ID NO. 6) was sequenced using techniques known in the art, and six mutations from the native sequence previously published for TEV were found. These changes were, according to numbering in Allison, *supra*, GC to CG at nucleotide 5768-5769, A to G at nucleotide 5773, A to G at nucleotide 6235, T to C at nucleotide 6314, and A to G at nucleotide 6961. The mutations were left unmodified as they did not affect the protease activity of NIa as shown by the results presented herein.

The cassette expression vectors presented herein, which exploit the proteolytic processing strategy of the TEV NIa protease, possess the advantages particular to the TEV NIa protease. First, NIa is a highly specific proteinase whose cleavage sequence has been well characterized (Carrington, *et al.*, 1988; Dougherty, *et al.*, 1988, *supra*; W. Dougherty, *et al., Virology*, 171:356-364, 1989; Dougherty, *et al., Virology*, 172:145-155, 1989). Second, NIa retains activity *in vitro* when cleavage sequences are inserted into several locations in TEV polyproteins (Carrington, *et al.*, 1988, *supra*; Dougherty, *et al.*, 1988, *supra*) or into non-viral proteins (Parks, *et al., J. Gen. Virol.*, 73:775-783, 1992). Finally,

NIa cleaves its substrate heptapeptide properly *in vivo* when expressed as a transgene in plants (Restrepo-Hartwig, *et al., J. Virology*, 66:5662-5666, 1992).

In one embodiment of the TEV-NIa-based expression cassette vectors provided herein, the NIa protease functions *in vitro* to cleave polypeptides containing inserted coding sequences for many different polypeptides ranging in size from 1 to as many as about 800 amino acids. In most of the constructs tested, cleavage was so effective that non-processed precursors could not be detected. In only two cases (an illustration is shown with pPRO1,SNG in Example 4) were minimal amounts of non-cleaved precursors detected, indicating a lack of complete processing. These *in vitro* results suggest utility of this approach for *in vivo* applications as well wherein the vectors are introduced into suitable plants by electroporation into plant protoplasts using methods well known in the art. (See for instance, *Current Protocols in Molecular Biology*, Ed. by F.M. Ausubel, Current Protocols, Vol. 1, §9.3.2-3, 1993). Transformed protoplasts can be harvested and grown into full transgenic plants (C. A. Rhodes, *et al., Science* 240:204-207, 1988).

In alternative embodiments, NIa-based expression cassette vectors are used in systems other than those involving plant cells. In general, the expression cassette of this invention can be used in any system in which the NIa protease has activity, for example, insect bacteria, mammalian, and other eukaryotic cells if operatively linked to suitable expression control elements such as a promoter, and a polyadenylation sequence, so as to bring about replication of the attached segment in a vector suitable for the type of cell line selected. However, for prokaryotic cells it may be necessary to reengineer the vector to bias it for codon specific organisms (see C.J Noren, *et al., Science*, 244:182, 1989). For example, as is well known, *Bacillus* spp. generally prefer more A/T rich nucleotide sequences.

The choice of vector to which a cassette of this invention is operatively linked depends directly, as is well known in the art, on the host cell to be transformed and the functional properties desired, e.g., vector replication and protein expression, these being limitations inherent in the art of constructing recombinant molecules. The vector itself may be of any suitable type, such as a viral vector (RNA or DNA), naked straight-chain or circular DNA, or a vesicle or envelope containing the nucleic acid material to be inserted into the cell. Techniques for construction of lipid vesicles, such as liposomes, are well known. Such liposomes may be targeted to particular cells using other conventional techniques, such as providing an antibody or other specific binding molecule on the exterior of the liposome (see, e.g., A. Huang, *et al., J. Biol. Chem.*, 255:8015-8018, 1980). In one embodiment of the invention, transient expression is contemplated wherein expression of the polypeptide is driven either by conventional transcriptional promoters or by plant viral vectors. In another embodiment, the TEV-NIa based cassette vector is used in prokaryotic systems since NIa proteases from different potyvirus have been shown to be active when expressed in bacterial cells (Garcia, *et al., Virology*, 170:362-369, 1989; Vance, *et al., Virology*, 191:19-30, 1992). The TEV NIa based expression vector can be advantageously used, therefore, whenever it is desirable to achieve equimolar production of two peptides in bacterial expression systems by inserting the NIa cassette into a bacterial expression vector, such as members of the pUC vector family. Other insect and animal cells known in the art to be useful in expression of recombinant proteins can also be used. For instance, the cassette vectors can be used in production of recombinant antibodies wherein it is desirable to achieve equimolar amounts of the heavy and light chains. In another embodiment, the cassette vectors provided herein are used to produce molecules that spontaneously assemble a two subunit complex, such as an enzyme. In yet another embodiment, a vector having more than two insertion sites is used to express multimers of any type.

Proteins expressed in the cassette vectors of this invention contain additional or extraneous amino acid residues at both N- and C-termini as a consequence of the NIa target heptapeptide and the cloning strategy used. The schematic diagram of Figure 1C illustrates the amino acid additions at N- and C-termini that result when in the proteins (open boxes) are cloned at either *Sma* I (Sm) or *Stu* I (ST) insertion sites of pPRO1. The amino acid represented by 'X' will depend on the restriction site used for cloning. In some cases one or more of the extraneous amino acids can be incorporated into the protein because it is already native to its sequence and would not have to be engineered in.

Due to the inclusion of additional amino acids at both termini of the cloned peptides, the biological activity of some proteins expressed in this system may be affected. However, one skilled in the art will know how to purify the produced proteins and treat them to clip off the extraneous residues. For instance, as shown in Figure 1C, the heterogenous proteins after cleavage by the protease can have among the extraneous terminal amino acids an undefined amino acid (represented by 'X') immediately next thereto at either end. If 'X' is selected to be a methionine and the produced peptide contains no other methionines, the peptide can readily be treated with cyanogen bromide to remove the extraneous residues. For example, the coat protein of TMV, which contains no methionines, can be expressed in one or both of the insertion sites, purified, and then can be treated with cyanogen bromide to provide the coat protein sequence free of extraneous terminal residues. One skilled in the art will be able to similarly utilize enzymes that cleave peptides between two particular residues to clip off the terminal extraneous residues from product heterogeneous peptides.

Several practical applications of the NIa cassette expression vectors utilizing its expression in plants as a transgene are also contemplated herein. For instance, coat protein mediated resistance (CPMR) to viral infections can generally be obtained only against viruses of the same taxonomic group as the one whose coat protein was used as

the vaccine (Fitchen & Beachy, *Annu. Rev. Microbiol.*, 47:739-763, 1993). To engineer coat protein mediated resistance (CPMR) against viruses that belong to different taxonomic groups, sequences encoding two or more viral coat proteins from different taxonomic groups can be inserted into insertion sites of a NIa-based vector having two or more insertion sites. Alternatively, an insect resistance gene can be combined with a virus resistance gene. In an alternative embodiment, the vector of this invention can be used to express a selectable marker plus any other gene encoding a protein of the size contemplated herein.

In yet another embodiment of this invention, described in full detail in U. S. Patent Application Serial No. 08/192,477 cofiled herewith, and incorporated herein by reference, the vector into which the cassette is ligated is a modification of the "infectious cDNA clone" of the tobacco mosaic virus to which is operably linked the promoter of the T7 polymerase. Highly infectious RNA transcripts of a full-length cDNA of the U1(common) strain of TMV have been produced *in vitro* using bacteriophage T7 RNA polymerase (Dawson, *et al., Proc. Natl. Acad. Sci USA*, 83:1832-1836, 1986; Meshi, *et al., Proc. Natl. Acad. Sci. USA*, 83:5043-5047, 1986). Alternatively, when inoculated into tobacco plants and other suitable host plants, this transcript causes systemic viral infection. Therefore, the vector of this invention can also be used to simultaneously provide systemic resistance to insect and virus in plants when inserted into the infectious cDNA clone of TMV.

In this embodiment of the invention, to accommodate the cassette to be inserted therein, the cDNA encoding the TMV movement protein is deleted from the TMV infectious clone, and the NIa-based cassette is ligated in its place, thereby creating a modified viral vector. Nucleotide sequences encoding heterologous peptides ligated into the insertion sites of the NIa-based cassette contained within the modified infectious clone can be inoculated into host plants for expression therein. Therefore, in this embodiment of the

invention the coat proteins of plant viruses belonging to a different taxonomic group than TMV, or other genes capable of protecting a plant against insect or disease, can be ligated into the insertion sites of the NIa-based cassette in the infectious clone vector for production in the host plant. Since the modified infectious clone vector retains the native gene encoding the coat protein of TMV, a cassette with two insertion sites can be used to express multiple CP sequences confer CPMR against viruses from three different taxonomic groups. If recombinant plants transformed with a gene encoding the wild type movement protein of the TMV, such as plant line 277 (Deom, *et al., Cell*, 69:221-224, 1992) are inoculated with the modified infectious clone vector, the viral infection will spread systemically. This modified infectious clone vector takes advantage of the extremely high level of expression characteristic of the viral system, and can be used to economically produce large amounts of polypeptides, virions suitable for use as vaccines, etc. One skilled in the art will appreciate that such product polypeptides and/or virions can be purified from plant leaves using standard methods (Bruening, *et al., Virology*, 71:498-517, 1976).

In initial experiments, constructs containing NIa and the CP of TMV (Figure 3A) were introduced in *Nicotiana tabacum* via *Agrobacterium tumefaciens* transformation. Preliminary data indicate that TMV CP expressed *in vivo* as part of pPRO1 confers CPMR (data not shown). Additional constructs with an insert that encodes a viral coat protein and a gene encoding β-glucuronidase will enable use of GUS activity as a probe for the levels of expression of the CP. Since the activity of the CP is destroyed if the protease does not cleave in the exact place anticipated, this experiment showed the specificty of the NIa protease for cleaving multiple exogenous peptides. This approach will be useful for studying those examples in which there is poor correlation between the levels of CP accumulation and the degree of plant viral resistance, providing additional important data on the molecular mechanism(s) of CPMR in these cases.

The following examples illustrate the manner in which the invention can be practiced. It is understood, however, that the examples are for the purpose of illustration and the invention is not to be regarded as limited to any of the specific materials or conditions therein.

EXAMPLE 1

CONSTRUCTION OF pPRO1 VECTORS

Recombinant DNA manipulation and *E. coli* transformation were carried out according to existing protocols (Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989). The DNA inserts used for the assembly of the different constructs were obtained by the polymerase chain reaction (PCR) using equipment and techniques provided by Perkin Elmer Cetus (Emoryville, CA). The sequences of primers used for amplification are detailed in Table 1, the prefix indicating the gene to which they are targeted.

The expression cassette vector pPRO1 (Figures 1A and 1B) was assembled in pBluescript II KS (+) (Stratagene, San Diego, CA) under the transcriptional control of a T7 promoter by directional insertion of PRO1 (SEQUENCE ID NO. 5) at the *Sac* I - *EcoR* I sites of the multiple cloning site, rendering pPRO1. NIa and 5'-non-translated (5-NTR) sequences from TEV were obtained by PCR using as DNA template a full length TEV cDNA clone (kindly provided by Dr. J. Carrington, Texas A&M University). Oligonucleotide primers for amplification of NIa were TEVNIA.N and TEVNIA.C (SEQUENCE ID NOS. 7 and 8, respectively). These two primers amplified the NIa open reading frame (Figure 1B) plus the sequences encoding the two specific heptapeptide cleavage sequences located at each end of NIa in the TEV genome and contained, in addition, either *Xba* I and *Sma* I (TEVNIA.N) or *Stu* I and *EcoR* I (TEVNIA.C) restriction enzyme sites. The PCR product was directionally inserted pBluescript using *Xba* I and *EcoR* I to yield vector

pBCNIa. Oligonucleotide primers used for PCR amplification of the 5'-NTR of TEV were TEVNTR.5 and TEVNTR.3 (SEQUENCE ID NOS. 9 and 10, respectively). These primers contained either *Sac* I and *Bgl* II (TEVNTR.5) or *Sma* I (TEVNTR.3) restriction enzyme cleavage sites. The final step in the assembly of pPRO1 was a *Sac* I-*Sma* I directed insertion of the TEV-5 NTR resulting from the PCR reaction into vector pBCNIa. Mutagenesis at the heptapeptides in the TEV sequence encoding the protease cleavage recognition sites was accomplished with primers TEVNIA.N2 and TEVNIA.C3 (SEQUENCE ID NOS. 11 and 12, respectively) which contained either one or two nucleotide changes (when compared to TEVNIA.N and TEVNIA.C, respectively) that mutated the glutamine located at position -1 (relative to the cleavage site) to histidine to introduce an *Nco* I insertion site useful for recovering the recombinant clones from the cloning vector pBCN1a.

The cDNAs for different open reading frames (ORFs) encoding heterogenous peptides inserted into pPRO1 included those encoding tobacco mosaic virus (TMV) and soybean mosaic virus (SMV) coat proteins (CP), as well as the *uidA* gene encoding the β-glucuronidase (GUS) activity from *E. coli.* These ORFs were obtained by PCR using as template publicly available nucleotide sequences. The nucleotide sequence of tobacco mosaic virus RNA, first published by P. Goelet, *et al. (Proc. Natl. Acad. Sci. U.S.A.*, 79:5818-5822, 1982) is publicly available from EMBL and Genebank databases under Accession Numbers V01408 and J02415. The nucleotide sequence of the CP gene of soybean mosaic virus, first published by A. Eggenberger, *et al., J. Gen. Virol.*, 70:1853-1860, 1989, is available from EMBL and Genebank databases under Accession Number D00507. The gene encoding GUS, first disclosed by R. A. Jefferson, *et al.*, (*Proc. Natl. Acad. Sci. U.S.A.*, 83:8447-8451, 1986) and available from EMBL and Genebank databases under Accession Number M14641, was obtained from Clontech. For PCR to obtain the ORF of TMV CP, primers TMV CP 51 (SEQUENCE ID NO. 13 was used at the 5' end and TMV CP 31 (SEQUENCE ID NO. 14) was used at the 3' end. For PCR

to obtain the ORF of SMV CP, primer SMV CP N1 (SEQUENCE ID NO. 15) was used at the 56' end and primer SMV CP C2 (SEQUENCE ID NO. 16) was used at the 3' end. For PCR to obtain the ORF of GUS, primer GUS N2 (SEQUENCE ID NO. 18) was used at the 5' end and primer GUS C1 (SEQUENCE ID NO. 19) was used at the 3' end.

TABLE 1

SEQUENCES OF THE OLIGONUCLEOTIDE PRIMERS USED

TEVNIA.N	5'-GCTCTAGA CCCGGG GAACCAGTCTATTTCCAAGGG-3'	(SEQ. ID NO. 7)
TEVNIA.C	5'-GCGAATTCAAGGCCT CCCTTGCGAGTACACCAATTCA-3'	(SEQ. ID NO. 8)
TEVNTR.5	5'-GCCGAGCTC AGATCT AAATAACAAATCTCAACACAACA-3'	(SEQ. ID NO. 9)
TEVNTR.3	5'-TCCCCCGGG CATGGCTATCGTTCGTAAATGG-3'	(SEQ. ID NO. 10)
TEVNIA.N2b	5'-TGGCCCGGG GAACCAGTCTATTTCCATGGG-3'	(SEQ. ID NO. 11)
TEVNIA.C3[b]	5'-GCGAATTCAAGGCCT CCCATGGGAGTACACCAATTCA-3	(SEQ. ID NO. 12)
TMVCP.51	5'-AAAGGCCT TCTTACAGTATCACTACTCC-3'	(SEQ. ID NO. 13)
TMVCP.31	5'-AGGCCCGGG AGTTGCAGGACCAGAGGTCC-3'	(SEQ. ID NO. 14)
SMVCP.N1	5'-AAAGGCCT TCAGGCAAGGAGAAGG-3'	(SEQ. ID NO. 15)
SMVCP.C2	5'-AGGCCCGGG CTGCGGTGGGCCCATGC-3'	(SEQ. ID NO. 16)
GUS.N2	5'-AAAGGCCT GTAGAAACCCCAACCCG-3'	(SEQ. ID NO. 17)
GUS.C1	5'-CGGAATTC TCATTGTTTGCCTCCCTGCTG-3'	(SEQ. ID NO. 18)

[a] Nucleotides annealing to the target genes are underlined with a single line, whereas nucleotides corresponding to the restriction enzyme recognition sequences are doubly underlined.

[b] Nucleotides changed in TEVNIA.N2 and TEVNIA.C3, when compared with TEVNIA.N. and TEVNIA.C, respectively, are marked by an asterisk underneath.

PCR products corresponding to SMV- and TMV-CP genes were digested with *Stu* I and *Sma* I and inserted either at the *Sma* I or the *Stu* I sites of pPRO1 (Figure 1), depending on the construct. The PCR product corresponding to the *uidA* ORF was digested with *Stu* I and *EcoR* I and inserted at the C terminus of NIa in pPRO1.

EXAMPLE 2

IN VITRO TRANSCRIPTION AND TRANSLATION

One µg of plasmid pPRO1 DNA containing the inserted heterologous ORFs purified from *E. coli* through QIAprep mini columns (Qiagen, Chatsworth, CA) was first linearized with *Sal* I (which cleaves downstream of pPRO1), and subsequently transcribed *in vitro* with T7 RNA polymerase (Epicentre Technologies, Madison, WI). Size and integrity of transcribed mRNA were confirmed by agarose gel electrophoresis. Approximately one µg of mRNA was used to program *in vitro* translation in 25 µL volume reactions using a nuclease treated rabbit reticulocyte lysate system (Promega, Madison, WI) according to the manufacturer's protocol. Proteins were synthesized in a nuclease treated rabbit reticulocyte lysate in the presence of ^{35}S-Met and then analyzed by SDS-PAGE (12.5% polyacrylamide) and autoradiography. However, since TMV CP contains no methionine residues, ^{3}H-Leu was used when the TMV CP ORF was translated *in vitro*. Proteins translated *in vitro* were analyzed by autoradiography following SDS-PAGE according to the method of U. Laemmli (*Nature*, [London] 227:680-685, 1970).

As shown in Figure 2, upon *in vitro* transcription and subsequent *in vitro* translation in the presence of ^{35}S-Met, pPRO1 gave the expected translated peptide of approximately 49 kDa. Experimental results demonstrate that this protein corresponded to NIa since it exhibited the proper proteolytic activity when expressed in pPRO1 as part of a polyprotein.

Other minor bands were also detected, some of which could be due to the autoproteolysis that releases the VPg (the protein linked to the 5' end of the viral RNA) from the protease domain in NIa during post-translational processing of TEV as described in W. Dougherty, *et al.* (*Virology*, 183:449-456, 1991).

Construction of Vectors Expressing TMV CP

To confirm that pPRO1 encodes NIa protease activity, several constructs were engineered in which the CP ORF from tobacco mosaic tobamovirus (TMV) was inserted into the cassette vector provided herein. These constructs are shown schematically in Figure 3A. The first two constructs, pPRO1.NT and pPRO1.TN, contained the TMV CP sequence in the C-terminal or N-terminal cloning sites, respectively. To demonstrate that processing of the resultant polyprotein was due to recognition and cleavage of the specific heptapeptides by the NIa protease and not to non-specific degradation, two additional controls were designed. First, the C-terminal NIa protease domain was removed with a frameshift mutation at the unique BamHI site, resulting in pPRO1TΔN (Figure 3A). In this construct, processing is not expected despite the presence of the naturally occurring cleavage sequence. Second, using methods described in Example 1, the two target heptapeptides were mutated to include a Gln to His change at the -1 position. This mutation at the cleavage site has been previously shown to inhibit the specific processing by NIa in TEV (Dougherty, *et al.*, 1988, *supra*; Dougherty, *et al.*, 1989, *supra*). The resulting mutant cassette vector was named pPRO4 and the corresponding pPRO4.NT and pPRO4.TN were also constructed as shown in Figure 3A.

In vitro transcription and translation of TMV CP-containing constructs in the above described rabbit reticulocyte lysate in the presence of ^{3}H-Leu, upon analysis by SDS-PAGE (15% polyacrylamide) and fluorography, revealed the expected patterns and sizes of labeled proteins as shown in Figure 3B. In addition to the 49 kDa protein, a band corresponding to a protein of approximately 18 kDa was detected in pPRO1.NT and

pPRO1.TN. 18 kDa is the expected size of TMV CP when expressed in pPRO1 constructs. The CP produced from pPRO1.TN was slightly larger than that produced from pPRO1.NT, in accordance with the numbers of amino acid residues added when the cDNA was cloned at the *Sma* I site versus the *Stu* I site (see Figure 1C). On the other hand, the major proteins resulting from constructs pPRO4.NT and pPRO4.TN migrated at positions corresponding to the size of the precursor polypeptide containing NIa plus TMV CP (68 kDa). Finally, when the protease domain from NIa was absent (pPRO1.T$_{\Delta}$N) a single protein of about 28 kDa, corresponding to the truncated protein, was detected.

Results of the *in vitro* translation followed by immunoprecipitation analyses of these vectors are shown in Figure 3C respectively. Immunoprecipitation assays were based upon previously described protocols with minor modifications. Briefly, 20 µL aliquots of *in vitro* translation reactions were diluted to 100 µL with TBSN (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% Nonidet P-40) and pre-incubated with protein A Sepharose beads (Sigma, St. Louis, MO) for 15 minutes on ice. After removing the beads, one µL was added of an appropriate dilution of a polyclonal antibody raised against TMV CP (ATCC# PVAS - 135) by standard techniques well known in the art . The mixture was incubated for 2-4 hours at 4°C with slow shaking. Subsequently, protein A Sepharose beads previously blocked with rabbit reticulocyte lysate were added and the mixture was kept on ice for 15 minutes with occasional shaking. The Sepharose beads were recovered and washed twice with 0.5 M LiCl, 20 mM Tris-HCl pH 8, once with TBSN, and once with H_2O. Finally, beads containing immunoprecipitated labeled proteins were resuspended in SDS-PAGE loading buffer and the proteins were analyzed as described above.

Immunoprecipitation reactions of the proteins produced *in vitro* using an anti-TMV CP antibody resulted in precipitation of the expected proteins (Figure 3C). Only those

peptides which included TMV CP sequences were selectively immunoprecipitated, whereas the 49 kDa NIa protein was not. These data clearly demonstrate that pPRO1 functions as predicted.

Several experiments were carried out to determine whether or not proteolytic processing could occur *in trans*. The labeled peptide that was translated from pPRO1.T$_{\Delta}$N was not processed when non-labeled 49 kDa protein translated from pPRO1 was used as source of NIa proteinase (data not shown). This result is in agreement with previously reported data. (J. Carrington and W. Dougherty, 1987, *supra)*.

EXAMPLE 3

PROTEOLYTIC PROCESSING OF TWO DIFFERENT PROTEINS INTRODUCED IN pPRO1

pPRO1 was further tested with the introduction of coding sequences for two different heterologous proteins into the two insertion sites. ORFs encoding coat proteins from viruses belonging to different groups, SMV (s; potyvirus) and TMV (T), were inserted to create constructions having the heterologous ORFs in the two possible positions. Figure 4A shows the resulting constructs pPRO1.SNT and pPRO1.TNS. As shown in Figure 4B, *in vitro* transcription and translation of these two constructs gave the predicted patterns of labeled proteins, resulting in the accumulation of proteins with the expected sizes of the NIa (49 kDa), SMV CP (around 30 kDa) and TMV CP (around 18 kDa). As expected, the coat proteins inserted at the *Sma* I site of pPRO1 gave slightly larger mature proteins than those inserted at the *Stu* I site due to incorporation of extra peptides as described in Figure 1C. Moreover, the more rapidly migrating proteins (predicted to be the TMV CP) co-migrated with proteins recovered following immunoprecipitation with anti-TMV CP antibody as in Example 2 above.

EXAMPLE 4

PROTEOLYTIC PROCESSING OF TWO OPEN READING FRAMES FROM UNRELATED PROTEINS

Another construct, pPRO1.SNG shown in Figure 5A, consisted of the SMV CP positioned at the *Sma* I insertion site of pPRO1 and the open reading frame encoding the β-glucuronidase activity (GUS) at the *Stu* I insertion site of pPRO1. As shown in Figure 5B, following *in vitro* translation in the presence of ^{35}S-Met, the expected profile of mature proteins was generated. The polypeptide synthesized upon translation of this construct has a predicted size of about 149 kDa, and is the largest that has been tested with the pPRO1 expression cassette. In this particular case, a high molecular weight band corresponding to a polypeptide of approximately 110 kDa was present in relatively low amounts. This protein probably corresponds to a fusion of the NIa and GUS peptides, implying that processing was not complete.

A time course *in vitro* translation reaction programmed with construct pPRO1.SNG and having samples withdrawn at the 5, 10, 15, 20, 30, 45, 60, and 90 minute intervals showed the predicted increase in the accumulation of the expected proteins with time as analyzed by SDS-PAGE (10% polyacrylamides) and autoradiography (Figure 5C). Even at short incubation times (15 min), no 149 kDa precursor could be detected, indicating efficient co-translational processing. However, pulse chase experiments with this construct did not demonstrate significant post translational processing of the low amounts of 110 kDa polypeptide (data not shown).

The foregoing description of the invention is exemplary for purposes of illustration and explanation. It should be understood that various modifications can be made without departing from the spirit and scope of the invention. Accordingly, the following claims are intended to be interpreted to embrace all such modifications.

SUMMARY OF SEQUENCES

Sequence ID No. 1 is an amino acid sequence for the consensus heptapeptide cleavage sequences that are cleaved by the NIa from TEV.

Sequence ID No. 2 is an amino acid sequence for the consensus heptapeptide cleavage sequences that are cleaved by the NIa from TEV.

Sequence ID No. 3 is an amino acid sequence for a self-recognized heptapeptide cleavage sequences at the N terminus of NIa in TEV.

Sequence ID No. 4 is an amino acid sequence for a self-recognized heptapeptide cleavage sequence C terminus of NIa in TEV.

Sequence ID No. 5 is a nucleotide sequence for PRO1 (Figure 1B).

Sequence ID No 6 is an amino acid sequence for PRO1 (Figure 1B).

Sequence ID No. 7 is a nucleotide sequence for a primer (TEVNIA.N) for amplification and cloning of cDNA encoding the nuclear inclusion a protein of tobacco etch potyvirus.

Sequence ID No 8 is a nucleotide sequence for a primer (TEVNIA.C) for amplification and cloning of cDNA encoding the nuclear inclusion a protein of tobacco etch potyvirus.

Sequence ID No. 9 is a nucleotide sequence for a primer (TEVNTR.5) for amplification and cloning of the 5' untranslated region of tobacco etch potyvirus.

Sequence ID No 10 is a nucleotide sequence for a primer (TEVNTR.3) for amplification and cloning of the 5' untranslated region of tobacco etch potyvirus.

Sequence ID No. 11 is a nucleotide sequence for a primer (TEVNIA.N2) for amplification and cloning of cDNA encoding the nuclear inclusion protein of tobacco etch potyvirus.

Sequence ID No 12 is a nucleotide sequence for a primer (TEVNIA.C3) for amplification and cloning of cDNA encoding the nuclear inclusion protein of tobacco etch potyvirus.

Sequence ID No. 13 is a nucleotide sequence for a primer (TMVCP.51) for amplification and cloning of cDNA encoding the tobacco mosaic virus coat protein.

Sequence ID No 14 is a nucleotide sequence for a primer (TMVCP.31) for amplification and cloning of cDNA encoding the tobacco mosaic virus coat protein.

Sequence ID No. 15 is a nucleotide sequence for a primer (SMVCP.N1) for amplification and cloning of cDNA encoding the soybean mosaic virus coat protein.

Sequence ID No. 16 is a nucleotide sequence for a primer (SMVCP.C2) for amplification and cloning of cDNA encoding the soybean mosaic virus coat protein.

Sequence ID No. 17 is a nucleotide sequence for a primer (GUS.N2) for amplification and cloning of cDNA encoding β-glucuronidase.

Sequence ID No. 18 is a nucleotide sequence for a primer (GUS.C1) for amplification and cloning of cDNA encoding β-glucuronidase.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: THE SCRIPPS RESEARCH INSTITUTE

(ii) TITLE OF INVENTION: A CASSETTE TO ACCUMULATE MULTIPLE PROTEINS THROUGH SYNTHESIS OF A SELF-PROCESSING POLYPEPTIDE

(iii) NUMBER OF SEQUENCES: 18

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Spensley Horn Jubas & Lubitz
(B) STREET: 1880 Century Park East, Suite 500
(C) CITY: Los Angeles
(D) STATE: California
(E) COUNTRY: USA
(F) ZIP: 90067

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: PCT
(B) FILING DATE: 03-FEB-1995
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Bostich, June M.
(B) REGISTRATION NUMBER: 31,238
(C) REFERENCE/DOCKET NUMBER: FD-3078

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (619) 455-5100
(B) TELEFAX: (610) 455-5110

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..7
(D) OTHER INFORMATION: /note= "where X appears, X can be any amino acid"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Glu Xaa Xaa Tyr Xaa Gln Gly
1 5

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..7
(D) OTHER INFORMATION: /note= "where X appears, X can be any amino acid"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Glu Xaa Xaa Tyr Xaa Gln Ser
1 5

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Glu Pro Val Tyr Phe Gln Gly
1 5

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:
(A) NAME/KEY: Peptide
(B) LOCATION: 1..7

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Glu Leu Val Tyr Ser Gln Gly
1 5

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1488 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:
(B) CLONE: PRO1

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 156..1481

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GAGCTCAGAT CTAAATAACA AATCTCAACA CAACATATAC AAAACAAACG AATCTCAAGC 60

AATCAAGCAT TCTACTTCTA TTGCAGCAAT TTAAATCATT TCTTTTAAAG CAAAAGCAAT 120

TTTCTGAAAA TTTTCACCAT TTACGAACGA TAGCC ATG CCC GGG GAA CCA GTC 173
Met Pro Gly Glu Pro Val
1 5

TAT TTC CAA GGG AAG AAG AAT CAG AAG CAC AAG CTT AAG ATG AGA GAG 221
Tyr Phe Gln Gly Lys Lys Asn Gln Lys His Lys Leu Lys Met Arg Glu
10 15 20

GCG CGT GGG GCT AGA GGG CAA TAT GAG GTT GCA GCG GAC GCA GGG GCG 269
Ala Arg Gly Ala Arg Gly Gln Tyr Glu Val Ala Ala Asp Ala Gly Ala
25 30 35

```
CTA GAA CAT TAC TTT GGA AGC GCA TAT AAT AAC AAA GGA AAG CGC AAG      317
Leu Glu His Tyr Phe Gly Ser Ala Tyr Asn Asn Lys Gly Lys Arg Lys
     40                  45                  50

GGC ACC ACG AGA GGA ATG GGT GCA AAG TCT CGG AAA TTC ATA AAC ATG      365
Gly Thr Thr Arg Gly Met Gly Ala Lys Ser Arg Lys Phe Ile Asn Met
 55                  60                  65                  70

TAT GGG TTT GAT CCA ACT GAT TTT TCA TAC ATT AGG TTT GTG GAT CCA      413
Tyr Gly Phe Asp Pro Thr Asp Phe Ser Tyr Ile Arg Phe Val Asp Pro
                 75                  80                  85

TTG ACA GGT CAC ACT ATT GAT GAG TCC ACA AAC GCA CCT ATT GAT TTA      461
Leu Thr Gly His Thr Ile Asp Glu Ser Thr Asn Ala Pro Ile Asp Leu
             90                  95                 100

GTG CAG CAT GAG TTT GGA AAG GTT AGA ACA CGC ATG TTA ATT GAC GAT      509
Val Gln His Glu Phe Gly Lys Val Arg Thr Arg Met Leu Ile Asp Asp
        105                 110                 115

GAG ATA GAG CCT CAA AGT CTT AGC ACC CAC ACC ACA ATC CAT GCT TAT      557
Glu Ile Glu Pro Gln Ser Leu Ser Thr His Thr Thr Ile His Ala Tyr
    120                 125                 130

TTG GTG AAT AGT GGC ACG AAG AAA GTT CTT AAG GTT GAT TTA ACA CCA      605
Leu Val Asn Ser Gly Thr Lys Lys Val Leu Lys Val Asp Leu Thr Pro
135                 140                 145                 150

CAC TCG TCG CTA CGT GCG AGT GAG AAA TCA ACA GCA ATA ATG GGA TTT      653
His Ser Ser Leu Arg Ala Ser Glu Lys Ser Thr Ala Ile Met Gly Phe
                155                 160                 165

CCT GAA AGG GAG AAT GAA TTG CGT CAA ACC GGC ATG GCA GTG CCA GTG      701
Pro Glu Arg Glu Asn Glu Leu Arg Gln Thr Gly Met Ala Val Pro Val
            170                 175                 180

GCT TAT GAT CAA TTG CCA CCA AAG AGT GAG GAC TTG ACG TTT GAA GGA      749
Ala Tyr Asp Gln Leu Pro Pro Lys Ser Glu Asp Leu Thr Phe Glu Gly
        185                 190                 195
```

```
GAA AGC TTG TTT AAG GGA CCA CGT GAT TAC AAC CCG ATA TCG AGC ACC      797
Glu Ser Leu Phe Lys Gly Pro Arg Asp Tyr Asn Pro Ile Ser Ser Thr
    200                 205                 210

ATT TGT CAC TTG ACG AAT GAA TCT GAT GGG CAC ACA ACA TCG TTG TAT      845
Ile Cys His Leu Thr Asn Glu Ser Asp Gly His Thr Thr Ser Leu Tyr
215                 220                 225                 230

GGT ATT GGA TTT GGT CCC TTC ATC ATT ACA AAC AAG CAC TTG TTT AGA      893
Gly Ile Gly Phe Gly Pro Phe Ile Ile Thr Asn Lys His Leu Phe Arg
                235                 240                 245

AGA AAT AAT GGA ACA CTG TTG GTC CAA TCA CTA CAT GGT GTA TTC AAG      941
Arg Asn Asn Gly Thr Leu Leu Val Gln Ser Leu His Gly Val Phe Lys
            250                 255                 260

GTC AAG AAC ACC ACG ACT TTG CAA CAA CAC CTC ATT GAT GGG AGG GAC      989
Val Lys Asn Thr Thr Thr Leu Gln Gln His Leu Ile Asp Gly Arg Asp
        265                 270                 275

ATG ATA ATT ATT CGC ATG CCT AAG GAT TTC CCA CCA TTC CCT CAA AAG      1037
Met Ile Ile Ile Arg Met Pro Lys Asp Phe Pro Pro Phe Pro Gln Lys
    280                 285                 290

CTG AAA TTT AGA GAG CCA CAA AGG GAA GAG CGC ATA TGT CTT GTG ACA      1085
Leu Lys Phe Arg Glu Pro Gln Arg Glu Glu Arg Ile Cys Leu Val Thr
295                 300                 305                 310

ACC AAC TTC CAA ACT AAG AGC ATG TCT AGC ATG GTG TCA GAC ACT AGT      1133
Thr Asn Phe Gln Thr Lys Ser Met Ser Ser Met Val Ser Asp Thr Ser
                315                 320                 325

TGC ACA TTC CCT TCA TCT GAT GGC ATA TTC TGG AAG CAT TGG ATT CAA      1181
Cys Thr Phe Pro Ser Ser Asp Gly Ile Phe Trp Lys His Trp Ile Gln
            330                 335                 340

ACC AAG GAT GGG CAG TGT GGC AGT CCA TTA GTA TCA ACT AGA GAT GGG      1229
Thr Lys Asp Gly Gln Cys Gly Ser Pro Leu Val Ser Thr Arg Asp Gly
        345                 350                 355
```

```
TTC ATT GTT GGT ATA CAC TCA GCA TCG AAT TTC ACC AAC ACA AAC AAT      1277
Phe Ile Val Gly Ile His Ser Ala Ser Asn Phe Thr Asn Thr Asn Asn
    360                 365                 370

TAT TTC ACA AGC GTG CCG AAA AAC TTC ATG GAA TTG TTG ACA AAT CAG      1325
Tyr Phe Thr Ser Val Pro Lys Asn Phe Met Glu Leu Leu Thr Asn Gln
375                 380                 385                 390

GAG GCG CAG CAG TGG GTT AGT GGT TGG CGA TTA AAT GCT GAC TCA GTA      1373
Glu Ala Gln Gln Trp Val Ser Gly Trp Arg Leu Asn Ala Asp Ser Val
                395                 400                 405

TTG TGG GGG GGC CAT AAA GTT TTC ATG AGC AAA CCT GAA GAG CCT TTT      1421
Leu Trp Gly Gly His Lys Val Phe Met Ser Lys Pro Glu Glu Pro Phe
            410                 415                 420

CAG CCA GTT AAG GAA GCG ACT CAA CTC ATG AGT GAA TTG GTG TAC TCG      1469
Gln Pro Val Lys Glu Ala Thr Gln Leu Met Ser Glu Leu Val Tyr Ser
        425                 430                 435

CAA GGG AGG CCT TGAATTC                                             1488
Gln Gly Arg Pro
    440
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 442 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Pro Gly Glu Pro Val Tyr Phe Gln Gly Lys Lys Asn Gln Lys His
  1               5                  10                  15

Lys Leu Lys Met Arg Glu Ala Arg Gly Ala Arg Gly Gln Tyr Glu Val
            20                  25                  30
```

Ala Ala Asp Ala Gly Ala Leu Glu His Tyr Phe Gly Ser Ala Tyr Asn
35 40 45

Asn Lys Gly Lys Arg Lys Gly Thr Thr Arg Gly Met Gly Ala Lys Ser
50 55 60

Arg Lys Phe Ile Asn Met Tyr Gly Phe Asp Pro Thr Asp Phe Ser Tyr
65 70 75 80

Ile Arg Phe Val Asp Pro Leu Thr Gly His Thr Ile Asp Glu Ser Thr
85 90 95

Asn Ala Pro Ile Asp Leu Val Gln His Glu Phe Gly Lys Val Arg Thr
100 105 110

Arg Met Leu Ile Asp Asp Glu Ile Glu Pro Gln Ser Leu Ser Thr His
115 120 125

Thr Thr Ile His Ala Tyr Leu Val Asn Ser Gly Thr Lys Lys Val Leu
130 135 140

Lys Val Asp Leu Thr Pro His Ser Ser Leu Arg Ala Ser Glu Lys Ser
145 150 155 160

Thr Ala Ile Met Gly Phe Pro Glu Arg Glu Asn Glu Leu Arg Gln Thr
165 170 175

Gly Met Ala Val Pro Val Ala Tyr Asp Gln Leu Pro Pro Lys Ser Glu
180 185 190

Asp Leu Thr Phe Glu Gly Glu Ser Leu Phe Lys Gly Pro Arg Asp Tyr
195 200 205

Asn Pro Ile Ser Ser Thr Ile Cys His Leu Thr Asn Glu Ser Asp Gly
210 215 220

His Thr Thr Ser Leu Tyr Gly Ile Gly Phe Gly Pro Phe Ile Ile Thr
225 230 235 240

Asn Lys His Leu Phe Arg Arg Asn Asn Gly Thr Leu Leu Val Gln Ser
245 250 255

Leu His Gly Val Phe Lys Val Lys Asn Thr Thr Thr Leu Gln Gln His
260 265 270

Leu Ile Asp Gly Arg Asp Met Ile Ile Ile Arg Met Pro Lys Asp Phe
275 280 285

Pro Pro Phe Pro Gln Lys Leu Lys Phe Arg Glu Pro Gln Arg Glu Glu
290 295 300

Arg Ile Cys Leu Val Thr Thr Asn Phe Gln Thr Lys Ser Met Ser Ser
305 310 315 320

Met Val Ser Asp Thr Ser Cys Thr Phe Pro Ser Ser Asp Gly Ile Phe
325 330 335

Trp Lys His Trp Ile Gln Thr Lys Asp Gly Gln Cys Gly Ser Pro Leu
340 345 350

Val Ser Thr Arg Asp Gly Phe Ile Val Gly Ile His Ser Ala Ser Asn
355 360 365

Phe Thr Asn Thr Asn Asn Tyr Phe Thr Ser Val Pro Lys Asn Phe Met
370 375 380

Glu Leu Leu Thr Asn Gln Glu Ala Gln Gln Trp Val Ser Gly Trp Arg
385 390 395 400

Leu Asn Ala Asp Ser Val Leu Trp Gly Gly His Lys Val Phe Met Ser
405 410 415

Lys Pro Glu Glu Pro Phe Gln Pro Val Lys Glu Ala Thr Gln Leu Met
420 425 430

Ser Glu Leu Val Tyr Ser Gln Gly Arg Pro
435 440

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:
(B) CLONE: TEVNIA.N

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCTCTAGACC CGGGGAACCA GTCTATTTCC AAGGG 35

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:
(B) CLONE: TEVNIA.C

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..37

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GCGAATTCAA GGCCTCCCTT GCGAGTACAC CAATTCA 37

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:
(B) CLONE: TEVNTR.5

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..38

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GCCGAGCTCA GATCTAAATA ACAAATCTCA ACACAACA 38

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:
(B) CLONE: TEVNTR.3

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..31

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TCCCCCGGGC ATGGCTATCG TTCGTAAATG G 31

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:
(B) CLONE: TEVNIA.N2b

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TGGCCCGGGG AACCAGTCTA TTTCCATGGG 30

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:
(B) CLONE: TEVNIA.C3b

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..37

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GCGAATTCAA GGCCTCCCAT GGGAGTACAC CAATTCA 37

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:
(B) CLONE: TMVCP.51

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..28

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AAAGGCCTTC TTACAGTATC ACTACTCC 28

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:
(B) CLONE: TMVCP.31

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: ..29

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AGGCCCGGGA GTTGCAGGAC CAGAGGTCC 29

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:
(B) CLONE: SMVCP.N1

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..24

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AAAGGCCTTC AGGCAAGGAG AAGG 24

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:
(B) CLONE: SMVCP.C2

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..26

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

AGGCCCGGGC TGCGGTGGGC CCATGC 26

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:
(B) CLONE: GUS.N2

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

AAAGGCCTGT AGAAACCCCA ACCCG 25

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vii) IMMEDIATE SOURCE:
(B) CLONE: GUS.C1

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..29

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CGGAATTCTC ATTGTTTGCC TCCCTGCTG 29

CLAIMS

1. An expression cassette comprising:

a nucleotide sequence encoding:

a) the nuclear inclusion (NIa) protease from tobacco etch virus;

b) multiple restriction endonuclease sites; and

c) self-cleavage sites for the protease, wherein the self-cleavage sites flank the protease and each restriction site, except at the termini of the nucleotide sequence.

2. An expression cassette vector comprising:

a) a nucleotide sequence encoding:

the nuclear inclusion (NIa) protease from tobacco etch virus;

multiple restriction endonuclease sites;

self-cleavage sites for the protease, wherein the self-cleavage sites flank the protease and each restriction site, except at the termini of the nucleotide sequence; and

b) expression control elements operably linked to the nucleotide sequence.

3. An expression cassette vector comprising:

a) a nucleotide sequence encoding:

the nuclear inclusion protein (NIa) from tobacco etch virus flanked by self-cleavage sequences therefor; and

restriction endonuclease sites flanking the self-cleavage sequences; and

b) expression control elements operably linked to the nucleotide sequence.

4. The vector of claim 2 wherein the nucleotide sequence further comprises:

a) an N-terminal start codon; and

b) a C-terminal stop codon.

5. The vector of claim 2 wherein at least one of the cleavage sequences encodes the amino acid sequence Sequence ID No. 1, wherein X is any amino acid.

6. The vector of claim 3 wherein at least one of the cleavage sequences encodes the amino acid sequence Sequence ID No. 2, wherein X is any amino acid.

7. The vector of claim 6 wherein the nucleotide sequence further comprises upstream of the open reading frames therein the 5' non-translated region from TEV RNA.

8. The vector of claim 2 wherein the N-terminus cleavage sequence encodes the amino acid sequence Sequence ID No. 4.

9. The vector of claim 8 wherein the C-terminus cleavage sequence encodes the amino acid sequence Sequence ID No. 5.

10. The vector of claim 2 wherein the restriction sites are blunt-ended.

11. The vector of claim 2 wherein the restriction sites are unique.

12. The cassette of claim 1 having the nucleotide sequence of Sequence ID No. 5.

13. The vector of claim 2 wherein one of the restriction endonuclease sites is a multiple restriction site.

14. The vector of claim 2 or 3 wherein a nucleotide sequence encoding a heterologous protein is inserted into each restriction endonuclease site.

15. An expression cell comprising the vector of claim 2.

16. An expression cell comprising the vector of claim 3.

17. The expression cell of claim 15 wherein the cell is a plant cell.

18. The expression cell of claim 15 wherein the cell is a prokaryotic cell.

19. A method for obtaining heterogeneous peptides in equimolar amounts comprising:

a) cleaving two or more the restriction endonuclease sites with enzymes specific therefor;

b) inserting DNA encoding a heterogeneous peptide into each cleaved restriction site;

c) transfecting a suitable cell with the vector;

d) culturing the transformed cell; and

e) obtaining the heterogeneous peptides in equimolar amounts.

20. The method of claim 19 wherein the cell is a plant cell.

21. The method of claim 20 wherein the plant cell is a plant protoplast and the culturing is *in vitro*.

22. The method of claim 19 wherein the cell is in a leaf of a plant and the culturing is *in vivo*.

23. The method of claim 19 wherein the cell is a prokaryote.

24. The vector of claim 2 or 3 wherein the promoter is the T7 polymerase promoter and the vector is derived from the infectious cDNA clone of TMV.

25. A plant cell infected with the vector of claim 24.

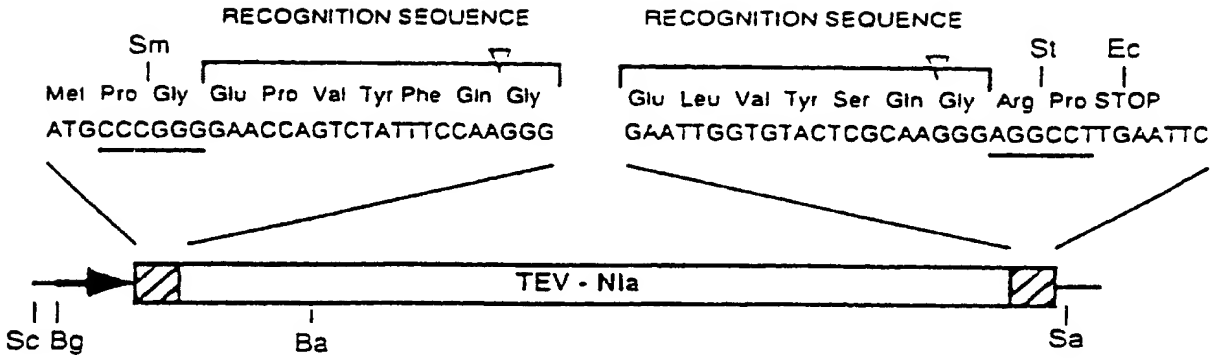


FIGURE 1A

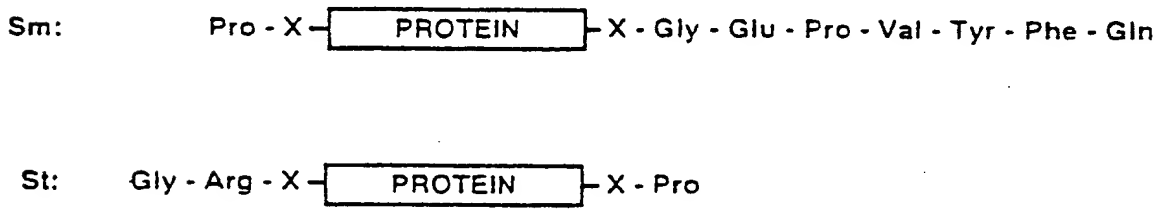


FIGURE 1C

Thursday, November 18, 1 3 11 PM
PRO1 Map2 (1 > 1488) Site and Sequence
Enzymes : 49 of 207 enzymes (Filtered)
Settings: Linear, Certain & Uncertain Sites, Standard Genetic Code

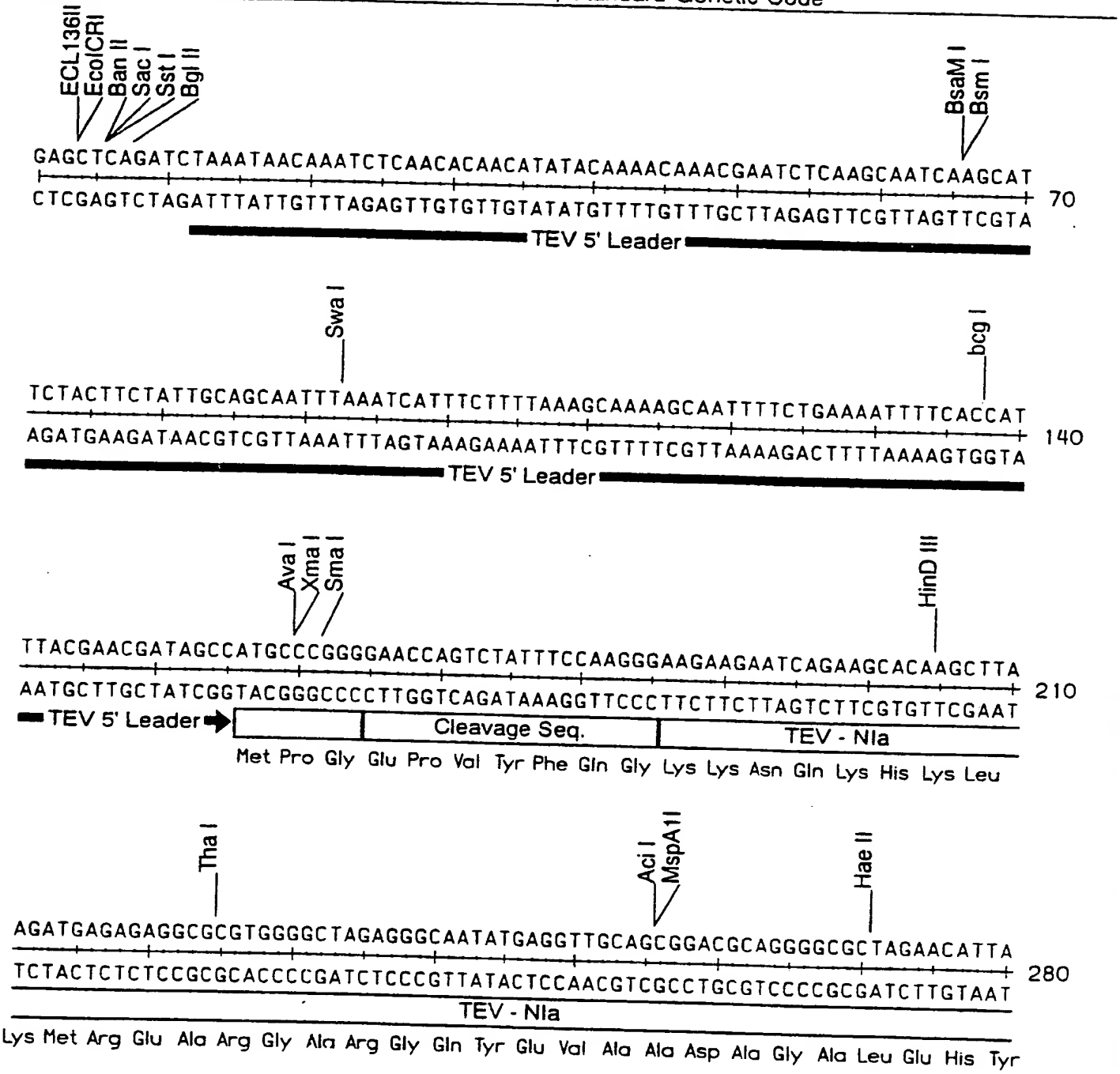
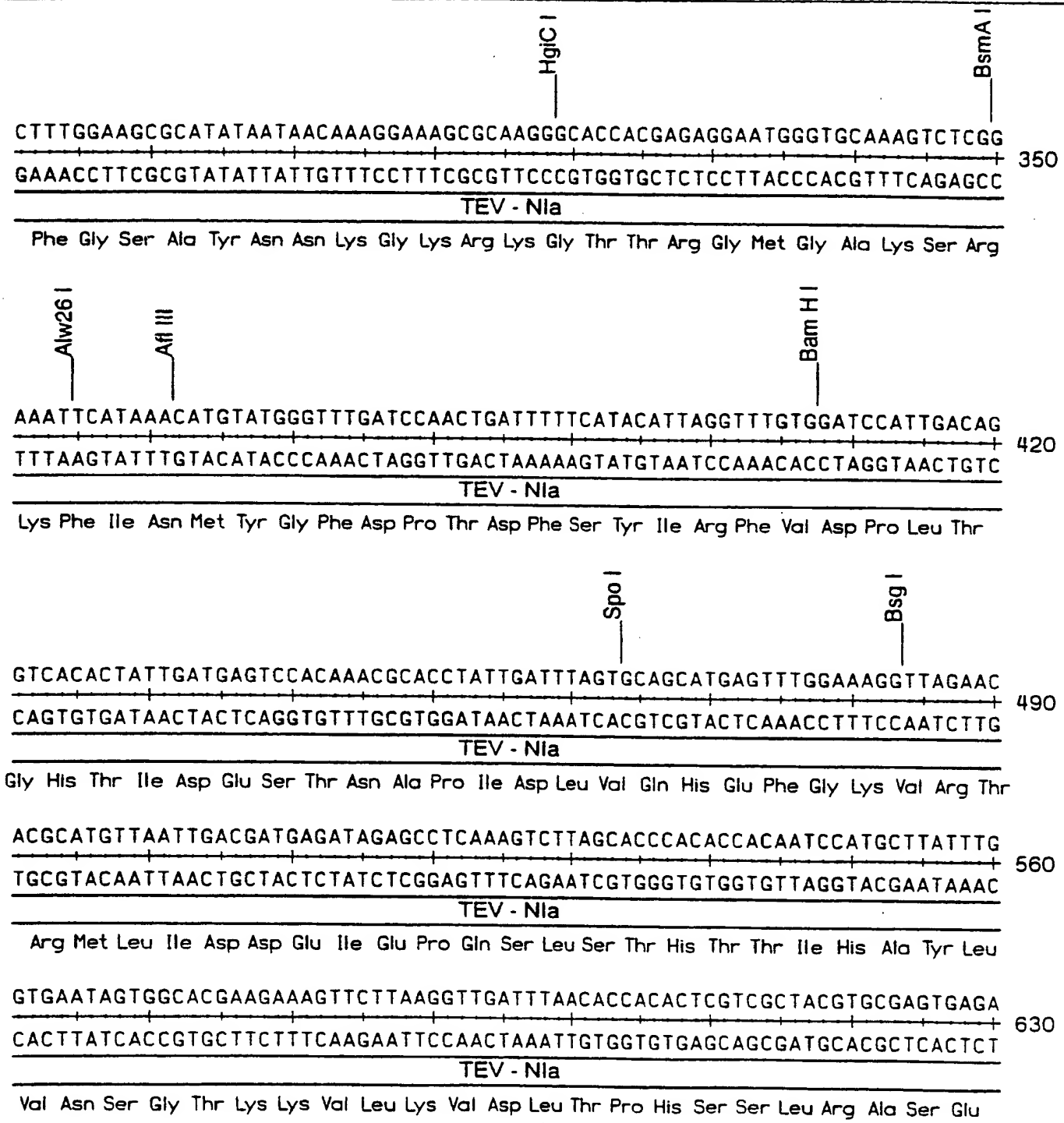
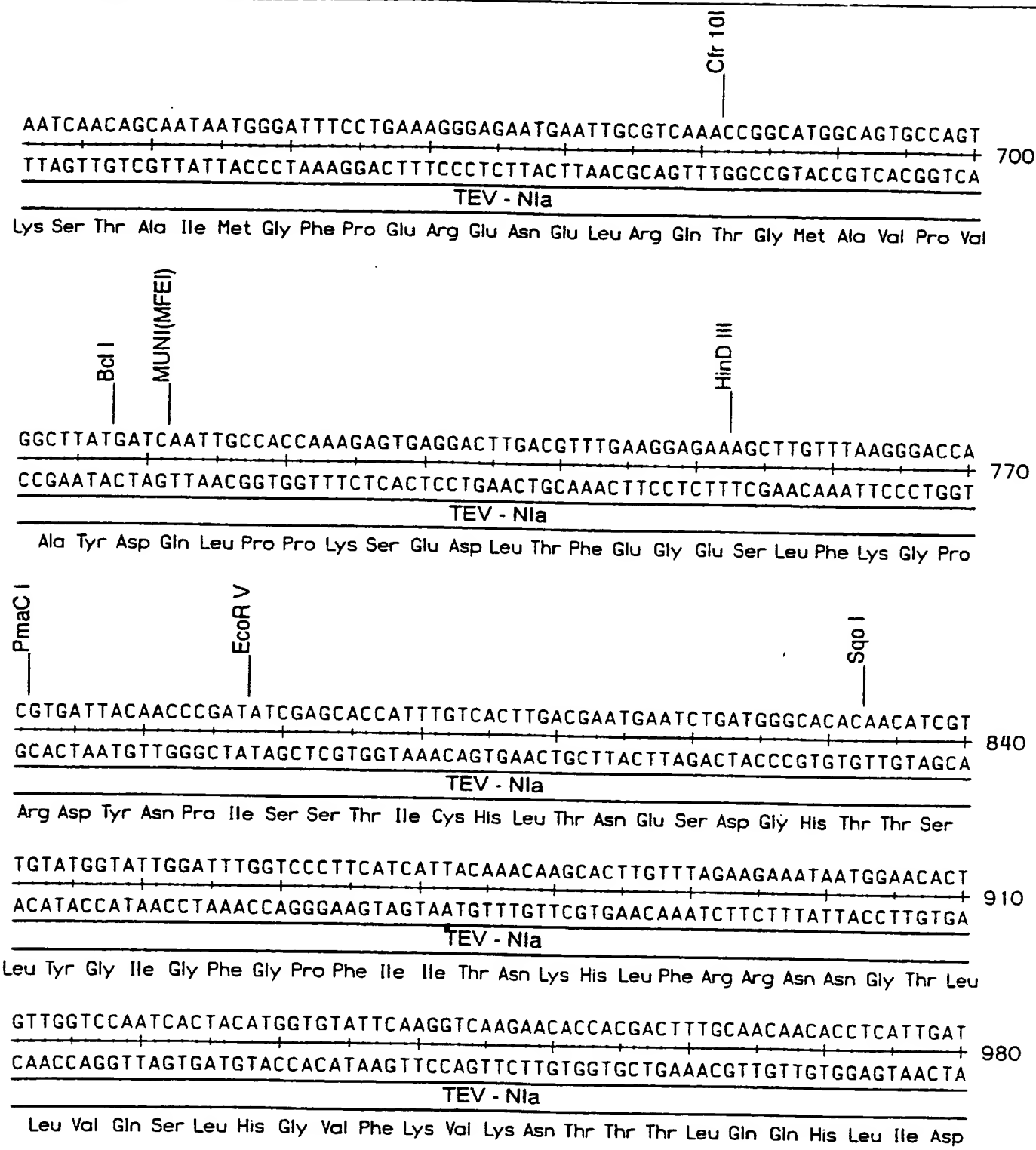
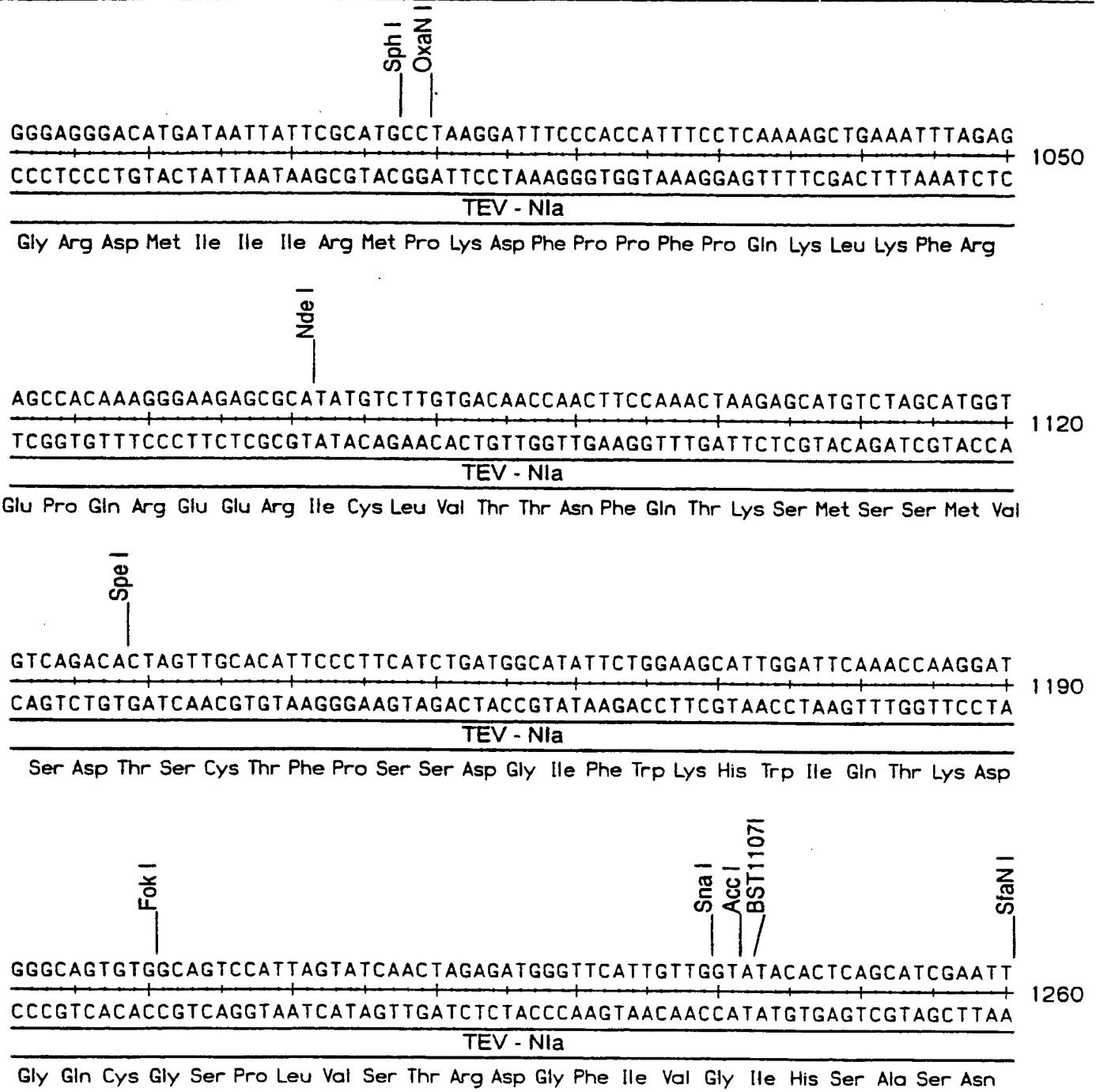


FIGURE 1B

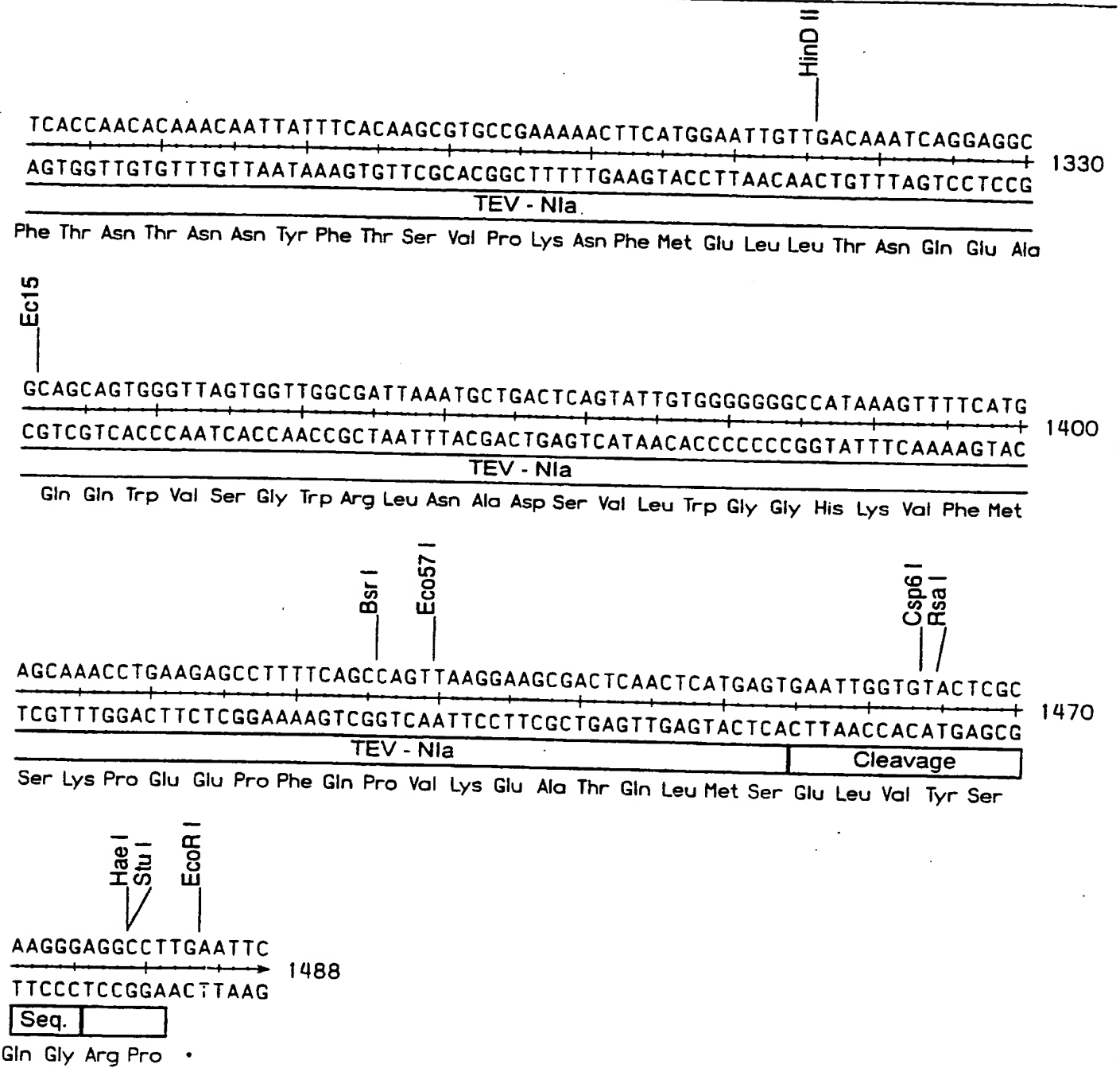
Thursday, November 18, 199_ 11 PM Page 2
PRO1 Map2 (1 > 1488) Site and Sequence




Cfr 10I
AATCAACAGCAATAATGGGATTTCCTGAAAGGGAGAATGAATTGCGTCAAACCGGCATGGCAGTGCCAGT
TTAGTTGTCGTTATTACCCTAAAGGACTTTCCCTCTTACTTAACGCAGTTTGGCCGTACCGTCACGGTCA
700
TEV - Nla
Lys Ser Thr Ala Ile Met Gly Phe Pro Glu Arg Glu Asn Glu Leu Arg Gln Thr Gly Met Ala Val Pro Val
Bcl I
MUNI(MFEI)
HinD III
GGCTTATGATCAATTGCCACCAAAGAGTGAGGACTTGACGTTTGAAGGAGAAAGCTTGTTTAAGGGACCA
CCGAATACTAGTTAACGGTGGTTTCTCACTCCTGAACTGCAAACTTCCTCTTTCGAACAAATTCCCTGGT
770
TEV - Nla
Ala Tyr Asp Gln Leu Pro Pro Lys Ser Glu Asp Leu Thr Phe Glu Gly Glu Ser Leu Phe Lys Gly Pro
PmaC I
EcoR V
Sqo I
CGTGATTACAACCCGATATCGAGCACCATTTGTCACTTGACGAATGAATCTGATGGGCACACAACATCGT
GCACTAATGTTGGGCTATAGCTCGTGGTAAACAGTGAACTGCTTACTTAGACTACCCGTGTGTTGTAGCA
840
TEV - Nla
Arg Asp Tyr Asn Pro Ile Ser Ser Thr Ile Cys His Leu Thr Asn Glu Ser Asp Gly His Thr Thr Ser
TGTATGGTATTGGATTTGGTCCCTTCATCATTACAAACAAGCACTTGTTTAGAAGAAATAATGGAACACT
ACATACCATAACCTAAACCAGGGAAGTAGTAATGTTTGTTCGTGAACAAATCTTCTTTATTACCTTGTGA
910
TEV - Nla
Leu Tyr Gly Ile Gly Phe Gly Pro Phe Ile Ile Thr Asn Lys His Leu Phe Arg Arg Asn Asn Gly Thr Leu
GTTGGTCCAATCACTACATGGTGTATTCAAGGTCAAGAACACCACGACTTTGCAACAACACCTCATTGAT
CAACCAGGTTAGTGATGTACCACATAAGTTCCAGTTCTTGTGGTGCTGAAACGTTGTTGTGGAGTAACTA
980
TEV - Nla
Leu Val Gln Ser Leu His Gly Val Phe Lys Val Lys Asn Thr Thr Thr Leu Gln Gln His Leu Ile Asp


Sph I
OxaN I
GGGAGGGACATGATAATTATTCGCATGCCTAAGGATTTCCCACCATTTCCTCAAAAGCTGAAATTTAGAG
1050
CCCTCCCTGTACTATTAATAAGCGTACGGATTCCTAAAGGGTGGTAAAGGAGTTTTCGACTTTAAATCTC
TEV - NIa
Gly Arg Asp Met Ile Ile Ile Arg Met Pro Lys Asp Phe Pro Pro Phe Pro Gln Lys Leu Lys Phe Arg
Nde I
AGCCACAAAGGGAAGAGCGCATATGTCTTGTGACAACCAACTTCCAAACTAAGAGCATGTCTAGCATGGT
1120
TCGGTGTTTCCCTTCTCGCGTATACAGAACACTGTTGGTTGAAGGTTTGATTCTCGTACAGATCGTACCA
TEV - NIa
Glu Pro Gln Arg Glu Glu Arg Ile Cys Leu Val Thr Thr Asn Phe Gln Thr Lys Ser Met Ser Ser Met Val
Spe I
GTCAGACACTAGTTGCACATTCCCTTCATCTGATGGCATATTCTGGAAGCATTGGATTCAAACCAAGGAT
1190
CAGTCTGTGATCAACGTGTAAGGGAAGTAGACTACCGTATAAGACCTTCGTAACCTAAGTTTGGTTCCTA
TEV - NIa
Ser Asp Thr Ser Cys Thr Phe Pro Ser Ser Asp Gly Ile Phe Trp Lys His Trp Ile Gln Thr Lys Asp
Fok I
Sna I
Acc I
BST1107I
SfaN I
GGGCAGTGTGGCAGTCCATTAGTATCAACTAGAGATGGGTTCATTGTTGGTATACACTCAGCATCGAATT
1260
CCCGTCACACCGTCAGGTAATCATAGTTGATCTCTACCCAAGTAACAACCATATGTGAGTCGTAGCTTAA
TEV - NIa
Gly Gln Cys Gly Ser Pro Leu Val Ser Thr Arg Asp Gly Phe Ile Val Gly Ile His Ser Ala Ser Asn

Thursday, November 18, 19 1 PM
PRO1 Map2 (1 > 1488) Site and Sequence



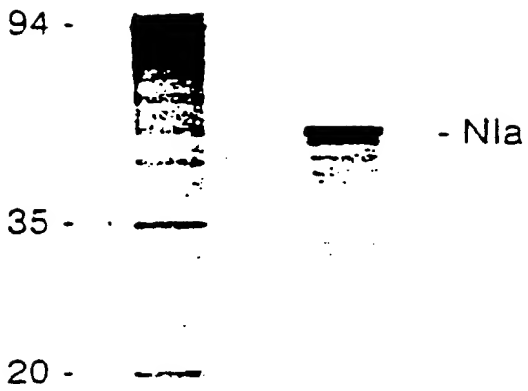
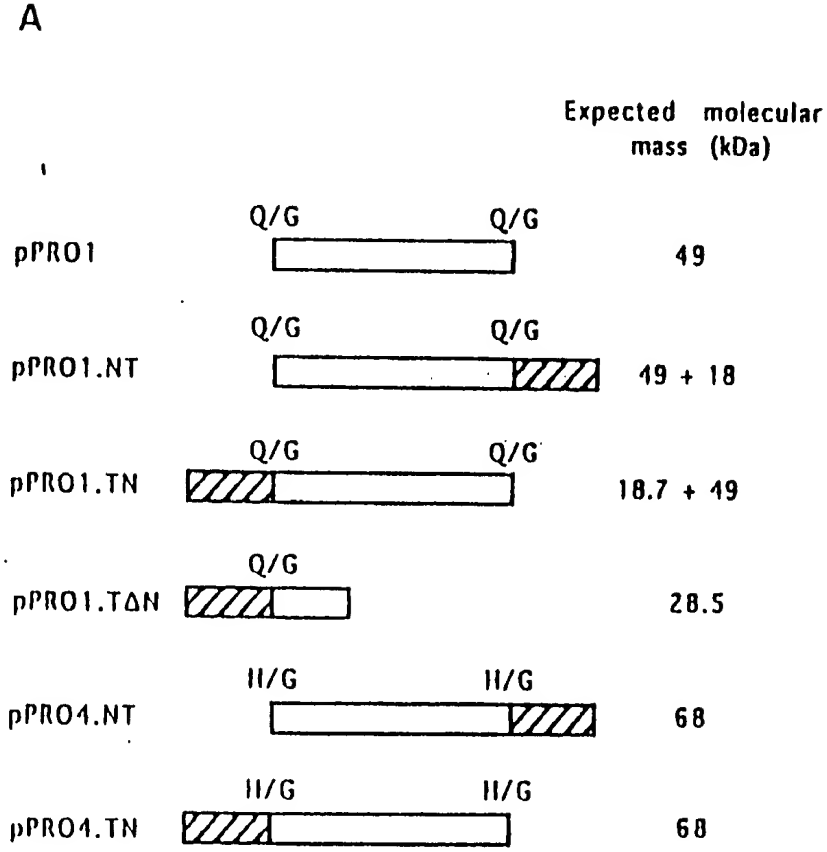
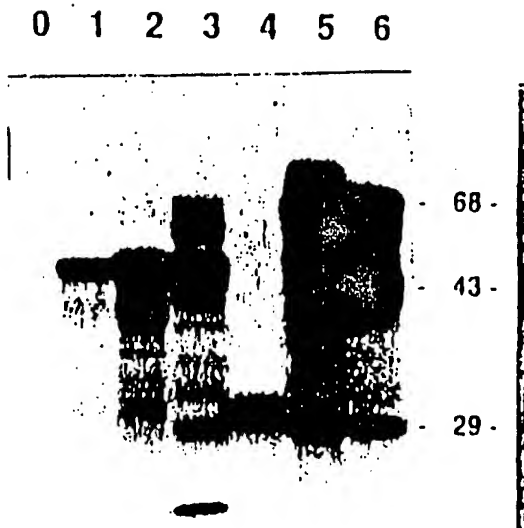


FIGURE 2





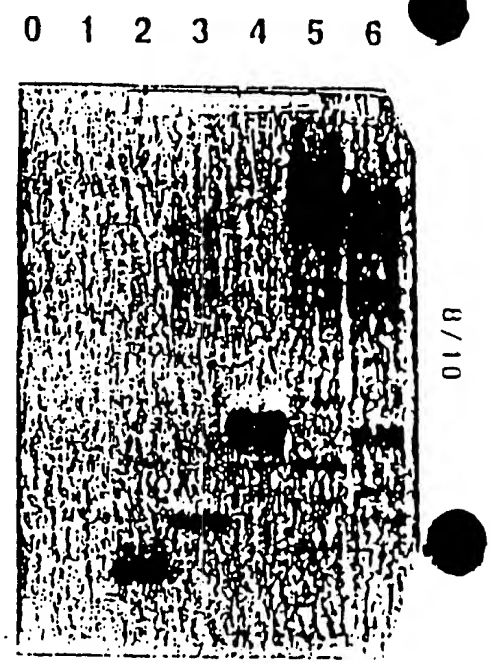
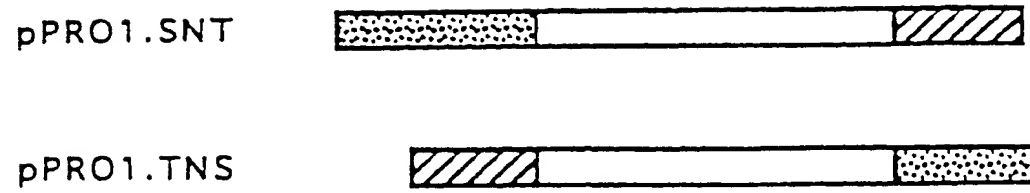


FIGURE 3

A



B

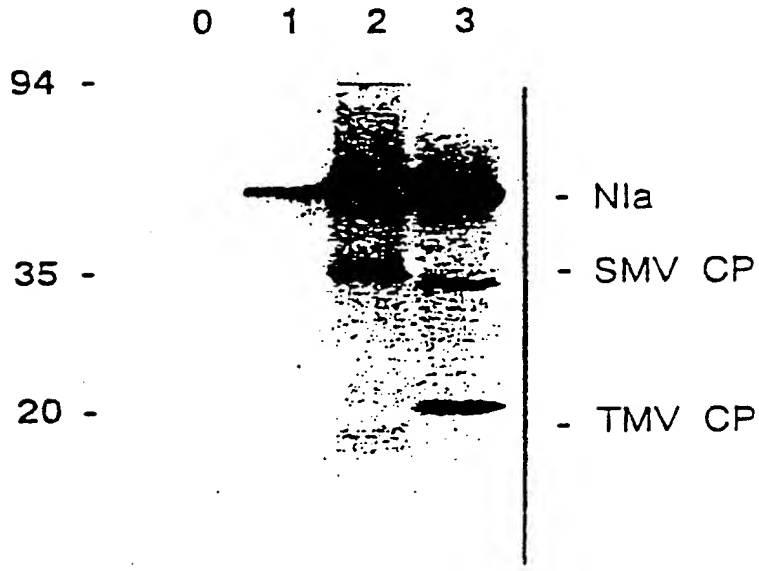
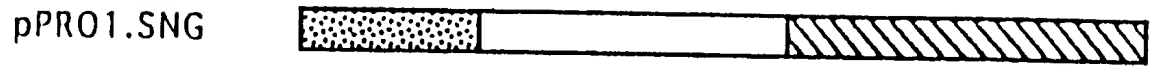
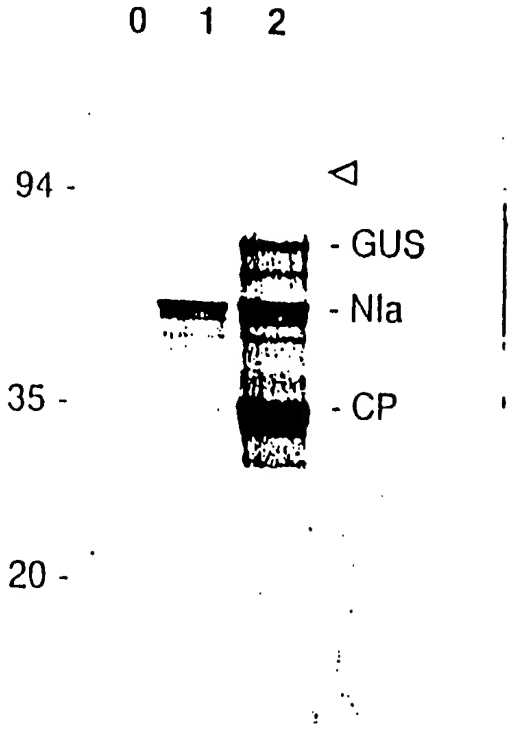


FIGURE 4

A.



B.



C.

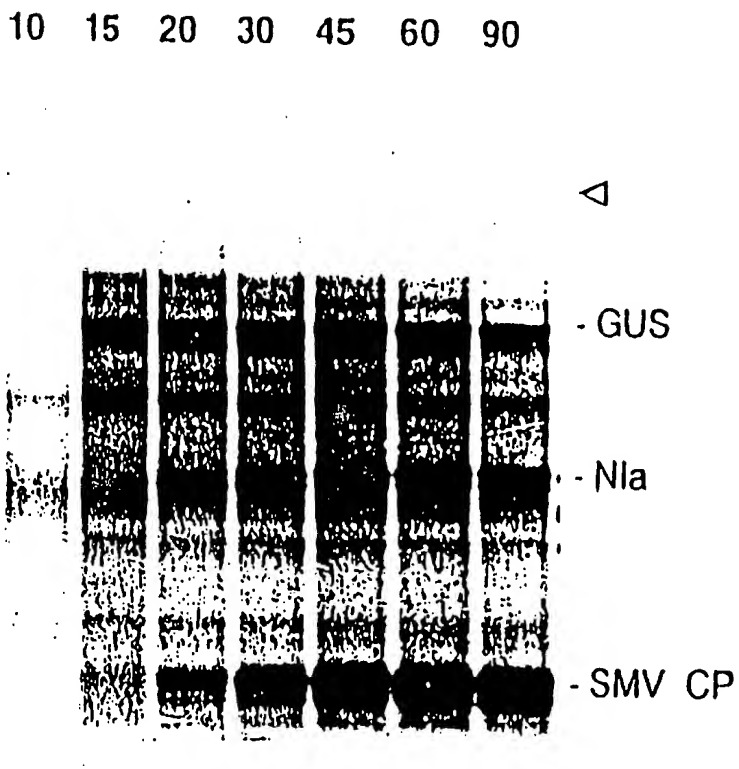


FIGURE 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/01495

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12N 15/00; C12P 21/00; A01H 1/04
US CL :435/69.1, 172.3, 240.4, 320.1; 800/205

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435 69.1, 172.3, 240.4, 320.1; 800/205

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, DIALOG

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US, A, 5,162,601 (SLIGHTOM) 10 NOVEMBER 1992, see columns 9-12 and 16-17.	1-25

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 04 APRIL 1995	Date of mailing of the international search report 18 APR 1995
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer ELIZABETH F. MCELWAIN Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)★